The National Environmental Information Exchange Network

Network Nodes at a Glance

Draft March 5, 2002

This document is intended to assist States that are applying for a National Environmental Information Exchange Network Grant (applications due April 1, 2002).   It is not intended and should not be used as a template for Grant application content or Node cost estimates.

Prepared by the Network Node Pilot Project – Beta Phase Group

Contents

Overview

State and EPA understanding of and experience with Network Nodes has developed substantially since 2000.   Under the sponsorship of the Interim Network Steering Group (INSG), several Node "How To" documents and specifications are now under development and should be available later this spring and summer.   This document summarizes the available high-level information about Nodes for Network Partners (Partners) applying for Network grants in the spring of 2002.

This document first provides a Node definition and explanation of Node functions (high-level). The majority focuses on helping States gauge the costs of establishing a Node.   It is unlikely that a "cookie cutter" template for Node costs will be available soon, if ever, because Node costs vary widely and are unique to each Partner's approach, technical requirements, level of existing technical expertise, and status of existing technical infrastructure.

The Network Steering Board (NSB) (which was created in February 2002 to succeed the INSG) is sponsoring teams to further develop existing Node Pilots, produce Node functional specifications, and to assess how standard software configuration templates (for use by Partners) might be developed.   For more information about the Node Pilot and related Network Node topics, see the section at the end of this document, "*Where to Find More Information*."

Definition of a Network Node

A **Network Node** *is a simple environmental information web service[1] that initiates or responds to requests for information, processes authorized queries, and sends/receives the requested information in the standard format: XML (eXtensible Markup Language).*

Nodes are defined by what they *do* (i.e., by their required functions), rather than what they are in a physical/hardware sense.   Several different hardware and software approaches and combinations may be used to establish a Node [e.g., Node functions performed using specialized Node software on a designated server (box) that serves no other purpose, Node functions performed using one or more types of software on more than one server, or Node functions performed within existing enterprise software on an existing server, etc.] Further, a Network premise is that Network participants will be free to choose their own approach to Node establishment – what is important is that the Node performs its functions.

Node Functions

Node functions enable Network information exchange and are divided into two categories: functions performed while requesting information (requestor Node functions), and functions

---

[1] A *web service* is software that exposes very simple functionality of business applications through the Internet.   Web services communicate with other web services via standards-based technologies that can be accessed by trading Partners independent of hardware, operating system, or programming environment.

performed while fulfilling information requests (requested Node functions).[2] Figure 1 depicts the high-level requestor Node functions.   Although the Node functional specifications will not be available until later in 2002,[3] Figure 2 provides the current understanding of detailed Node functions.

Figure 1.   High-Level (Requested) Node Functions
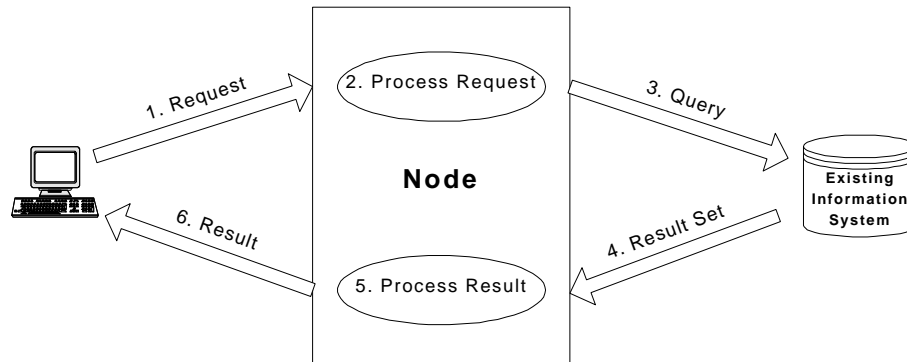
Figure 2. Detailed Node Functions

---
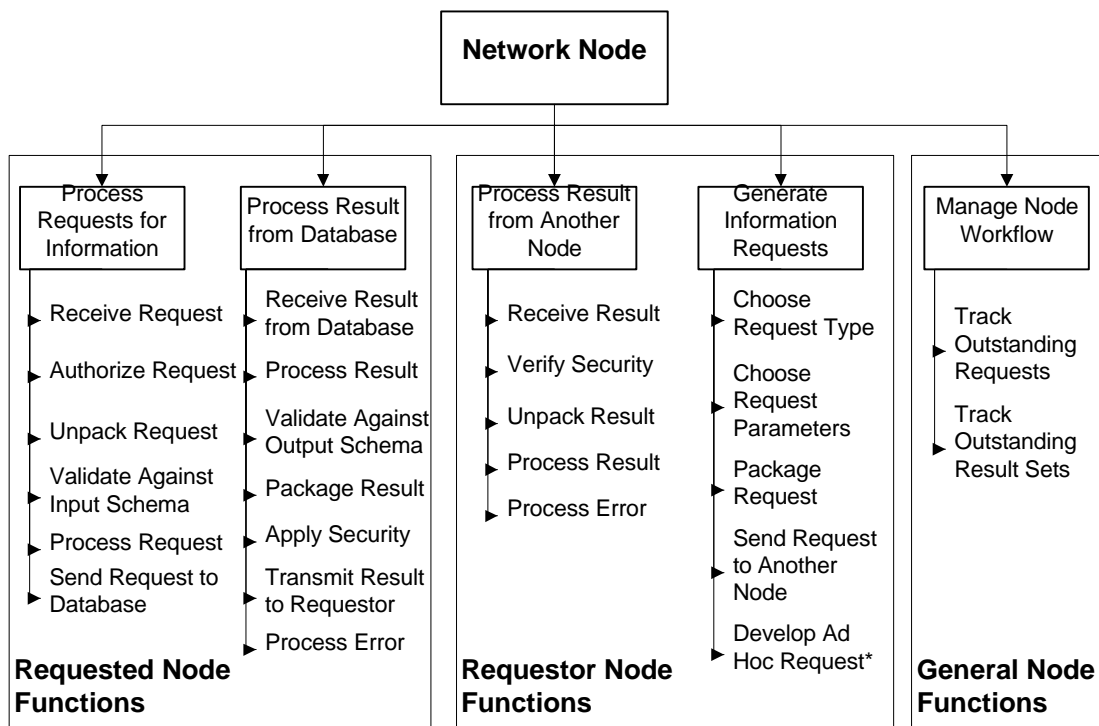
[2] To date, the Network Node Pilot Project – Beta Phase has focused on requestor Node functions.
[3] However, a general draft of expected Node functions will be available (by late March/early April) in the *Network Node Pilot Project – Beta Phase: Results Report.*

How Nodes Relate to the Grants

The FY 2002 Network Grants are intended to advance a Partner's readiness to participate in the Network. It is important to consider Nodes when applying for any of the FY 2002 Network grants, because it is impossible to participate in the Network without a Node.

The Network Readiness Grant requires each applicant to provide a proposal that addresses their commitment to participate in the Network, and the actual development of a Node or portal on the Network. While the One Stop Grant does not explicitly call for Node establishment, it is important that grant applicants consider Node development in their planning. For example, if a grant recipient is planning major infrastructure investments with their grant monies, such as an enterprise application integration, they should consider how easy it will be to build a Node once the new system is in place, and how easy it will be to connect information to the Node. Challenge Grants do not explicitly call for a Node, but may indirectly require a Node to meet the expectations of advancing the Network.

Node Costs

Partners preparing their grant applications probably care most about the question, "How much will my Node cost?" The following section provides information so that each Partner can roughly estimate the investments they will need to establish a Node. However, for the following four reasons, the enclosed costs information should not be used as a template for Node cost planning or the exact content of the Grant applications:

1. There are many approaches to building a Node. As standards such as SOAP and

---

Highlight 1

Node Costs: A *General* Breakdown

(The Costs for each Node will differ. The following information is *not* a Node cost template.)

Node Hardware and Software

- Server(s)*
- Firewall^
- Additional (IT) Network Components
    - E.g., cables
- Application Server^
- Middleware*
    - Vendor
    - Edition
    - Number and type of licenses
    - Upgrades
    - Service/Maintenance Agreements
    - Etc.
- Database Connectivity
- Security^

Node Operation

- Initial/First Establishment *
- Consulting*
- Staff Time
    - Gap Analysis/Tool Selection*
    - Maintenance
    - Troubleshooting*

Data Flows (Per Flow)

- Mapping database to Node*
- Query Development (i.e., SQL, Stored Procedures)*

Training

- XML*
- Web Services*
- Software (i.e., Middleware, Enterprise Application Integration)*

* = Costs likely to be much higher for initial Node establishment (and accompanying first Flow)

^ = Most State agencies already have this item or regularly incur costs related to this item.

XML are built into more and more new and existing applications, these options expand.

2. Different Partners will have different requirements for their Nodes in terms of scale and performance, and Partners will have different installed software bases.   In many cases, Partners will be able to use existing (or planned) software for most or all Node functions.

3. Partners will vary in how much on-staff expertise with the relevant technologies is available, and how much will have to be developed or out-sourced.

4. What costs will be "for the Node" (e.g., purchase of a dedicated box) versus for a given Flow or application will vary from Partner to Partner. Some Partners may dedicate their entire grants to purchase and configure a high performance application server and associated hardware, which would host their Node along with many other applications. Another Partner may elect to use open-source software (e.g., Linux-based) and connect it to an existing application. There is no way to usefully compare these two Node costs.

Partners may find it useful to use recent experience (redundant) with deploying web-based public access (read only) applications that draw on existing databases. Such projects share nearly all of the software and management issues associated with Node development.   Highlight 1 (above) identifies probable cost areas and factors.   Some costs, such as data quality, data completeness, and modification of existing system, may also impact the cost of sharing information through a Node.

Selecting the Node "Toolset" and Establishing the Node

Many tools and combinations of tools (hardware and software/middleware) can be used to establish a Node.   Tools have different functional options and range in both complexity and price.   Many of these options are likely to go far beyond the basic Network Node functional specifications,[4] and therefore may or may not be useful to the implementing agency.   Additional factors when considering specific tools are:

* The configurations and types of a Partner's existing information systems

* The participant's future information management needs and plans   (If a Partner is planning on a significant infrastructure upgrade, the selection of the Node tool must reflect a Partner's information technology strategy.)

* What, if anything, the Partner may want to use the Node for, besides Network participation

* The Partner's technical capacity.   If the participant does not have in-house expertise in Node technologies, and does not want to rely on outside expertise to establish a Node, then tools with user-friendly "wizards" and automated functions may be best.[5]

These factors and their consideration are similar to those relevant to other enterprise software choices (e.g., security or web server) that Partners make on a routine basis.

---

[4] Node functional specifications are expected by mid-late 2002.
[5] Note that some custom coding is likely to be needed for all Nodes.

Once the appropriate Node tool has been selected, Partners have to install their Node.   Node installation requires preliminary planning, technical expertise, and coordination. Technical expertise is necessary, as the Node must interface with the Network and the implementing agency's existing information system.   Typically, this is done in a similar fashion to implementation of an ODBC solution, with the added complexity of integration with firewall services. Without coordination with other program or IT staff, for instance security or firewall staff, Node installation will become difficult, if not impossible.   As Partners already know from their application integration efforts, these management and coordination costs can be significant. Finally, when planning Node implementation and to simplify troubleshooting, Beta Phase participants found it beneficial to establish Nodes in small, incremental steps.

Establishing Each Network Flow

Each Partner is expected to have only one Node. Like a firewall or web server, the Node will be used by many Flows and/or applications. Using the Node for a first Network Flow, regardless of the Flow's complexity or size, is likely to require more time (and therefore cost) due simply to Node for each additional Flow is likely to cost less than the first, although each Flow will have a unique set of information management needs, such as data quality, data completeness, any additional database or middleware modifications necessary for a successful flow, and assigning responsibility for each of these items.   As Partners gain (technical and administrative) experience with Flows, each subsequent Flow is likely to be more cost efficient.   It is important to note, however, that data quality, data completeness, and database modifications are not direct Node costs.

Node Costs: The Node Pilot Project Experience

A logical question when gauging Node costs is, "What did the Nodes that have already been established cost?" While both this question and its answer are timely, the costs of the Pilot Nodes do not accurately represent future Node costs for the following reasons: 1) the Node Pilot was a group "learning" effort for learning about Nodes and informing future Node establishment, rather than taking existing knowledge and applying it, 2) the Pilot States used only a few tools that are not the most expensive available (e.g., an evaluation version of Microsoft's Biztalk server), 3) the Pilot Project included pilot facility flows rather than Network Flows, 4) lessons learned from the Pilot Project are likely to reduce the cost of subsequent Node implementations, and 5) the Pilot had outside support provided by EPA.   Keeping in mind these cost caveats, the Pilot Node direct costs can be broken down per Node as follows:

- Up to $15k for Node tools
- Up to 80 hours of staff time (for Node implementation)
- Up to 40 hours of contractor time (excluding project coordination and conference calls)

Indirect costs are not included in the Pilot Node's cost estimates.   Indirect costs include tool selection/gap analysis, data quality and completeness, and identifying business requirements. Staff time and contractor time will vary depending on technical expertise.

The table below provides the high-level software architecture of the Nodes developed in the Beta Phase.

Table 1. Software Architecture of the Beta Nodes
Note: See the Beta Phase report for more detail.

Node Component	Microsoft -Based	Oracle -Based	XAware -Based
SOAP Listener	MS Internet Information Server 5.0 with ASAPI Listener	Jserv Server (Java Servlet)	Jakarta Tomcat
SOAP Processor	MSXML 4.0	SOAP Server (Java Servlet)	XAware
Object Handler	COM +	Oracle 9i AS	XAware
Data Mapper	MS BizTalk 2000	Oracle 9i AS with Apache Jserv	XAware
XML Processor	MS BizTalk 2000	Oracle 9i AS	XAware
DB Connectivity	Depends on existing information system and current methods for accessing databases. Examples include: Oracle ODBC, ADO, JDBC, Oracle Native Connection Driver, or SQL ODBC.		

Node Costs: Hypothetical Examples

Table 2 contains three hypothetical Node approaches, the assumptions in each approach, and the **estimated ranges** in cost for each approach.   **These are general hypothetical Node costs.** While based on current information, Table 2 (and Highlight 1) should not be used as a proxy for an actual cost assessment.   **Actual Node costs will vary.**

Table 2. General Hypothetical Node Costs
(Initial establishment and extra staff hours for one initial Flow)

Approach #1	Approach #2	Approach #3
- Node hardware must be purchased. - Non-enterprise low-price solution. - Assume you have all necessary accompanying software and hardware. - No additional staff training. - Consultants only retained for troubleshooting during implementation. - Existing information system and data are simple, and staff is capable of mapping to information system.	- Node hardware must be purchased. - Enterprise mid-price solution. - Assume you have all the necessary accompanying software and hardware. - Some additional staff training necessary. - Consultants retained for implementation only. - Existing information system and data are simple, and staff is capable of mapping to information system.	- Node hardware must be purchased. - High-price enterprise solution. - Some additional staff training necessary. - Consultants retained for implementation and maintenance and operation. - Existing information system and data are complex, and staff is capable of mapping to information system (increase in staff hours).

Node Costs		Node Costs		Node Costs	
Hardware	$3,000-$20,000	Hardware	$3,000-$20,000	Hardware	$3,000 - $20,000
Software	$0 - $6,000	Software	$7,000-$17,000	Software	$20,000-$35,000
Consulting	$0 - $5,000	Consulting	$10,000-$20,000	Consulting	$20,000-$30,000
Training	$0	Training	$5,000-$10,000	Training	$15,000-$25,000
Other Direct & Indirect Costs*	$1,000-$25,000	Other Direct & Indirect Costs*	$1,000-$25,000	Other Direct & Indirect Costs*	$1,000-$25,000
Staff Hours**	$8,000-$15,000	Staff Hours*	$5,000-$12,000	Staff Hours*	$7,000-$14,000
Total	**$12,000-$71,500**	**Total**	**$31,000-$104,000**	**Total**	**$66,000-$149,000**

* These costs might include, but are not limited to, backup, disaster recovery, and maintenance.
 ** Includes extra time for initiating one initial Flow. Note that Flow-specific costs are likely to vary widely.

Towards Technology Templates

Given that Nodes can be developed using standard software, and that many Partners use similar software (e.g., Microsoft and Oracle), many Partners have suggested the idea of "technology   ble 1 for an example of an early high-level template.) Technology templates are specific combinations of software (and eventually the code used in conjunction with those tools for Network purposes) that are available for Network participants to use "off t build their Nodes.   The NSB intends to sponsor the creation of these templates based on experience and Partner interest.   Some Partners may wish to estimate their Node development costs now, but delay deployment until the Node Specification and these templates (if a demand for them has been established and they are developed) are available.   Other Partners, especially those with established applications already using Node technologies, will probably not be interested in templates and will implement their Nodes immediately. Only a few tools out of the dozens available have been tested; as such it is too early to develop specific Node template recommendations.

Where to Find More Information

Below is contact information, information on grant guidance, and reference material on the Pilot Project and the Network.   The final Beta Phase report will be completed in late March, and will be available on the WISER and ECOS websites.

Grant Guidance:
- Grant website, www.epa.gov/neengprg.

Network Nodes and Network Node Pilot Projects:
- Beta Phase: visit WISER's Node Beta conference room, Network Node Pilot Project Beta Phase, http://www.ecos.org/wiser/conferences/ (Password protected)
  Contacts:
    - o Dennis Burling, Nebraska DEQ, Beta Phase Co-Chair – dennis.burling@ndeq.state.ne.us.
    - o Connie Dwyer, US EPA, Beta Phase Co-Chair – dwyer.connie@epa.gov.
    - o Anna Brooks, Ross & Associates, Beta Phase Coordinator – anna.brooks@ross-assoc.com.
- *Network Nodes: A Primer* (December 14, 2001 draft) – on Grant website (see above)
- This document – on Grant website (see above), WISER's Node Beta Conference Room (see above), and ECOS, www.sso.org/ecos/

General Information on the Network:
- *Blueprint for a National Environmental Exchange Network (Blueprint)*: http://www.sso.org/ecos/eie/COMPLETE_BLUEPRINT_JUNE_01_FINAL.pdf.
- *Network Implementation Plan*: is available at the Grant website and ECOS website.